US007620628B2

(12) **United States Patent**
Kapur et al.

(10) **Patent No.: US 7,620,628 B2**
(45) **Date of Patent: Nov. 17, 2009**

(54) **SEARCH PROCESSING WITH AUTOMATIC CATEGORIZATION OF QUERIES**

(75) Inventors: **Shyam Kapur**, Sunnyvale, CA (US); **Jignashu Parikh**, Jamnagar (IN); **Deepa Joshi**, Santa Clara, CA (US)

(73) Assignee: **Yahoo! Inc.**, Sunnyvale, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 378 days.

(21) Appl. No.: **11/006,466**

(22) Filed: **Dec. 6, 2004**

(65) **Prior Publication Data**

US 2006/0122979 A1 Jun. 8, 2006

(51) **Int. Cl.**
G06F 17/30 (2006.01)

(52) **U.S. Cl.** **707/4**; 707/6

(58) **Field of Classification Search** 707/3, 707/6, 100, 104.1, 4

See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

6,424,973 B1 7/2002 Baclawski

(Continued)

FOREIGN PATENT DOCUMENTS

WO WO 02/19147 A1 3/2002

OTHER PUBLICATIONS

Ipeirotis et al., "Probe, count, and classify: categorizing hidden web databases", Proceedings of the 2001 ACM SIGMOD international conference on Management of data , p. 67-78, May 2001. Retrived from the Internet: < URL: http://portal.acm.org/ft_gateway.cfm?id=375671&type=pdf&coll=ACM&dl=ACM&CFID=48417025&CFTOKEN=19109216>.*

(Continued)

Primary Examiner—John E Breene
Assistant Examiner—Dennis Myint
(74) *Attorney, Agent, or Firm*—Hickman, Palermo, Truong & Becker LLP

(57) **ABSTRACT**

Search results are processed using search requests, including analyzing received queries in order to provide a more sophisticated understanding of the information being sought. A concept network is generated from a set of queries by parsing the queries into units and defining various relationships between the units. From these concept networks, queries can be automatically categorized into categories, or more generally, can be associated with one or more nodes of a taxonomy. The categorization can be used to alter the search results or the presentation of the results to the user. As an example of alterations of search results or presentation, the presentation might include a list of "suggestions" for related search query terms. As other examples, the corpus searched might vary depending on the category or the ordering or selection of the results to present to the user might vary depending on the category. Categorization might be done using a learned set of query-node pairs where a pair maps a particular query to a particular node in the taxonomy. The learned set might be initialized from a manual indication of which queries go with which nodes and enhanced has more searches are performed. One method of enhancement involves tracking post-query click activity to identify how a category estimate of a query might have varied from an actual category for the query as evidenced by the category of the post-query click activity, e.g., a particular hits of the search results that the user selected following the query. Another method involved determining relationships between units in the form of clusters and using clustering to modify the query-node pairs.

5 Claims, 8 Drawing Sheets